The genetic analysis of repeated measures II. The Karhunen-Loève expansion

Molenaar, P.C.M.; Boomsma, D.I.

published in
Behavior Genetics
1987

DOI (link to publisher)
10.1007/BF01065503

document version
Publisher's PDF, also known as Version of record

Behavior Genetics, Vol. 17, No. 3, 1987

The Genetic Analysis of Repeated Measures. II The Karhunen-Loève Expansion

Peter C. M. Molenaar[1] **and Dorret I. Boomsma**[2]

Received 22 July 1985—Final 10 May 1986

A new approach to the genetic analysis of time series of arbitrary length and with arbitrary covariance function is outlined. This approach is based on the simultaneous eigenvalue decomposition of the covariance matrices of the original time series obtained from monozygotic (MZ) and dizygotic (DZ) twins. The method is illustrated with computer-simulated twin data.

KEY WORDS: time series; eigenvalue decomposition; genetic correlations; environmental correlations; twin data.

INTRODUCTION

A genetic analysis of time series, i.e., long stretches of repeated observations such as typically encountered in psychophysiological research, raises problems that are related to the proper handling of autocorrelation. For instance, a standard univariate technique such as ANOVA of repeated measures is based on the assumption of compound symmetry of the autocorrelation function. This means that observations at different time points t and t' should always have the same correlation irrespective of the lag $t' - t$. Box (1954) indicated that even moderate deviations from compound symmetry in an ANOVA of repeated measures lead to great distortions in probability levels for comparisons between time points.

In general, the autocorrelation of a time series will be some decreasing function of the lag $t' - t$ (cf. Box and Jenkins, 1976) and consequently will lack compound symmetry. Such lag-dependent autocorrelations are

[1] Department of Psychology, University of Amsterdam, Weesperplein 8, 1018 XA Amsterdam, The Netherlands.
[2] Department of Experimental Psychology, Free University, De Boelelaan 1115, 1081 HV Amsterdam, The Netherlands.

0001-8244/87/0500-0229$05.00/0

regularly found with psychophysiological time series (Lutzenberger *et al.*, 1980; Wastell, 1981). One therefore would like to have an alternative to the ANOVA of repeated measures that enables a robust genetic analysis of time series with arbitrary autocorrelation function. Preferably, such an alternative approach should identify the genetic and environmental autocorrelation structures underlying an observed time series and enable a complete description of the latent pattern of time-dependent genetic and environmental processes.

An approach that is consistent with these aims of dynamic genetic analysis involves the use of MANOVA in combination with simplex-type analysis (Boomsma and Molenaar, 1987). However, this approach is not practically feasible with lengthy (psychophysiological) time series. Consequently, an alternative genetic analysis accommodating time series of arbitrary length is required. In the following, such an alternative approach, based on the Karhunen-Loève expansion (cf. Ahmed and Rao, 1975), is outlined. The Karhunen-Loève expansion involves the decomposition of a time series into uncorrelated projections on the eigenvectors of the autocorrelation function. Stated otherwise, the time series is transformed into a sequence of uncorrelated variables, thus enabling the use of standard univariate techniques. Moreover, the Karhunen-Loève expansion applies in situations where the number of repeated measures exceeds the number of subjects. In such cases, the resulting covariance matrix of observations is singular, but its decomposition into genetic and environmental components may still be achieved using the proposed method.[3]

In the following sections a basic genetic model for arbitrary, i.e., stationary or nonstationary, time series is presented. Next, the dynamic genetic analysis based upon the Karhunen-Loève expansion is discussed in some detail and illustrated by means of several applications to simulated data. In the closing section we consider several generalizations of the proposed analysis, particularly those related to spectral analysis.

DEFINITIONS

A univariate time series $y(t)$ can be conceived of as a member of an ensemble of time-dependent functions which are generated by some random scheme (Brillinger, 1975). The mean function of $y(t)$ is defined by

$$\text{ave}\,[y(t)] = m_y(t),$$

where the average is taken over members of the ensemble of random

[3] We wish to thank an unknown reviewer for pointing this out.

functions at each time t. Similarly, the covariance function of $y(t)$ is defined by

$$\text{cov}\,[y(t), y(t')] = c_y(t, t').$$

According to these definitions both the mean function and the covariance function are time-varying or nonstationary.

A time series $y(t)$ can be stationary in several respects. For instance, $y(t)$ can have a stationary mean function,

$$m_y(t) = m_y,$$

or a stationary covariance function,

$$c(t, t') = c(0, t' - t) = c(u),$$

or both. Notice that a stationary covariance function depends only on the length of the interval $u = t' - t$ and therefore is invariant under a translation along the time axis.

A GENETIC MODEL FOR TIME SERIES

Consider the following basic genetic model for an observed time series $y(t)$:

$$y(t) = G(t) + E(t),$$

where $G(t)$ and $E(t)$ are latent time series of genetic and nonshared environmental influences, respectively, and where $G(t)$ and $E(t)$ are mutually uncorrelated. The covariance functions of $y(t)$, $G(t)$, and $E(t)$ are denoted by $c_y(t, t')$, $c_g(t, t')$, and $c_e(t, t')$, respectively. For the moment, it is convenient (although not necessary) to assume that the corresponding mean functions are stationary, i.e., $m_y(t) = m_g(t) = m_e(t) = m$, where $m = 0$. With these provisions, let $\mathbf{y} = [y(1), \ldots, y(T)]'$ be a $T \times 1$ vector denoting $y(t)$ at a finite interval of times $t = 1, \ldots, T$. The $T \times T$ covariance matrix of $\mathbf{y}$ is

$$\mathbf{S}_y = \{c_y(t, t'); t, t' = 1, \ldots, T\}.$$

Defining $\mathbf{S}_g$ and $\mathbf{S}_e$ likewise, it then follows that

$$\mathbf{S}_y = \mathbf{S}_g + \mathbf{S}_e.$$

A GENETIC ANALYSIS OF TIME SERIES

In this section we consider a transformation of the original time series $y(t)$ into a sequence of uncorrelated variables, which can be analyzed

independently from each other by means of standard univariate techniques. The uncorrelated variables in question are similar to component scores as obtained with principal-components analysis and, accordingly, can be conceived of as projections of $y(t)$ on the principal components or eigenvectors of $\mathbf{S}_y$. A description of $y(t)$ in terms of a linear combination involving uncorrelated projections on the eigenvectors of $\mathbf{S}_y$ is called the Karhunen-Loève expansion of $y(t)$.

The required transformation of the original time series $y(t)$ into a sequence of uncorrelated variables is obtained from the eigenvalue decomposition of $\mathbf{S}_y$:

$$\mathbf{S}_y = \mathbf{PDP}',$$

where $\mathbf{D} = \text{diag}\,[d_1, \ldots, d_T]$ is the $T \times T$ diagonal matrix of eigenvalues and $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_T]$ is the associated $T \times T$ matrix of orthogonal eigenvectors. Accordingly,

$$\mathbf{y}^* = \mathbf{P}'\mathbf{y} = [y^*_1, \ldots, y^*_T]'$$

is a $T \times 1$ vector of uncorrelated variables y^*_i:

$$\text{cov}\,[y^*_i, y^*_j] = \mathbf{p}_i' \mathbf{S}_y \mathbf{p}_j = 0 \qquad \text{if } i \neq j.$$

Hence, a finite sample of a time series $y(t)$ with arbitrary covariance function can be transformed into a sequence of uncorrelated variables y^*_i, i = 1, . . . , T. The inverse transformation $y(t)$ may be obtained by

$$y(t) = \sum_{i=1}^{T} y^*_i p_i(t)$$

or, equivalently,

$$\mathbf{y} = \mathbf{Py}^*.$$

This inverse transformation is commonly called the Karhunen-Loève expansion and is exploited in the analysis of the basic genetic model.

Consider a sample of $4N$ finite sample paths of $y(t)$ obtained with N monozygotic (MZ) and N dizygotic (DZ) twin pairs, i.e.,

$$\begin{aligned} y_{kmn}(t) \qquad & t = 1, \ldots, T, \\ & k = 1 \text{ (MZ) or } 2 \text{ (DZ)}, \\ & m = 1, 2 \text{ (members of a pair)}, \\ & n = 1, \ldots, N \text{ (number of pairs)}. \end{aligned}$$

From the usual assumption that $\mathbf{S}_{y_{kmn}} = \mathbf{S}_y$, for all k, m, n, and that $m_{y_{kmn}} = m_y = 0$, it follows that the estimator of the covariance function of y(t) is

$$\hat{c}_y(t, t') = \sum_k \sum_m \sum_n y_{kmn}(t) y_{kmn}(t')/4N, \qquad t, t' = 1, \ldots, T.$$

The eigenvalue decomposition of $\hat{\mathbf{S}}_y = \{\hat{c}_y(t, t'); t, t' = 1, \ldots, T\}$ yields a $T \times T$ matrix $\hat{\mathbf{P}}$ of orthogonal eigenvectors $\hat{\mathbf{p}}_i$, $i = 1, \ldots, T$. Consecutively, the $4N$ finite sample paths $y_{kmn}(t)$ are transformed into $4N$ sequences of mutually uncorrelated variables

$$y^*_{i,kmn} = \hat{\mathbf{p}}_i' \mathbf{y}_{kmn}, \qquad i = 1, \ldots, T,$$

where $\mathbf{y}_{kmn} = [y_{kmn}(1), \ldots, y_{kmn}(T)]'$. For a given eigenvector $\hat{\mathbf{p}}_i$, then, we obtain $4N$ variables $y^*_{i,kmn}$, which can be analyzed independently from the remaining variables $y^*_{j,kmn}$ corresponding to eigenvectors $\hat{\mathbf{p}}_j, j \neq i$.

In a nutshell, the genetic analysis of $4N$ time series $y_{kmn}(t)$, $t = 1, \ldots, T$, has been transformed into T independent genetic analyses of $4N$ variables $y^*_{i,kmn}$. As the latter variables do not depend upon time t any more, each of the ensuing genetic analyses can be carried out by means of standard univariate techniques. Specifically, for $i = 1, \ldots, T$ maximum-likelihood estimates of the proportions of genetic and environmental variance based upon the mean squares of $\{y^*_{i,kmn}\}$ between and within monozygotic and dizygotic twin pairs may be obtained by

$$\text{EMS} = [\sigma_g, \sigma_e] = \begin{bmatrix} \Psi & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_g \\ \sigma_e \end{bmatrix},$$

where

$$\Psi = 2 \quad \text{for MSB (MZ)},$$

$$\Psi = 0 \quad \text{for MSW (MZ)},$$

$$\Psi = 1.5 \text{ for MSB (DZ)},$$

$$\Psi = 0.5 \text{ for MSW (DZ)}.$$

The likelihood function of these structural equations for the mean squares can be numerically optimized by means of standard methods. Thus, we obtain estimates of the proportion of genetic variance $[h_i^2 = \sigma^2_{g_i}/(\sigma^2_{g_i} + \sigma^2_{e_i})]$ and environmental variance $[e_i^2 = \sigma^2_{e_i}/(\sigma^2_{g_i} + \sigma^2_{e_i})]$ associated with each eigenvector $\hat{\mathbf{p}}_i$. Notice that at this point in the analysis any plausible genetic model for $y_i^*(t)$ can be tested. For instance, the presented approach can accommodate a model including genetic, common-family environmental, and within-family environmental components.

From the eigenvalue decomposition of a covariance matrix $\mathbf{S}_y$, it follows that

$$\text{var}\,[y^*_{i,kmn}] = \mathbf{p}_i'\mathbf{S}_y\mathbf{p}_i = d_i, \qquad \text{and}$$

$$\text{trace}\,\mathbf{S}_y = \sum_{t=1}^{T} c_y(t, t) = \sum_{i=1}^{T} d_i.$$

Hence, the proportion of genetic variance in the total variance of $y(t)$ is estimated by

$$\hat{h}^2 = \sum_{i=1}^{T} \hat{h}_i^2 \hat{d}_i.$$

In addition, the following estimates of the covariance functions of the genetic series $G(t)$ and the environmental series $E(t)$ are obtained:

$$\hat{\mathbf{S}}_g = \hat{\mathbf{P}}\mathbf{D}_g\hat{\mathbf{P}}', \qquad \hat{\mathbf{D}}_g = \text{diag}\,[\hat{h}_1^2\hat{d}_1, \ldots, \hat{h}_T^2\hat{d}_T],$$

$$\hat{\mathbf{S}}_e = \hat{\mathbf{P}}\hat{\mathbf{D}}_e\hat{\mathbf{P}}', \qquad \hat{\mathbf{D}}_e = \text{diag}\,[\hat{e}_1^2\hat{d}_1, \ldots, \hat{e}_T^2\hat{d}_T].$$

The estimates $\hat{h}^2$, $\hat{e}^2$, $\hat{\mathbf{S}}_g$, and $\hat{\mathbf{S}}_e$ pertain to the original time series $y(t)$. The corresponding estimators combine results obtained with transforms y^*_i and, consequently, can be conceived of as inverse transformations back to the original data.

Summarizing, we have presented a general approach to the genetic analysis of time series with arbitarary, i.e., possibly nonstationary, covariance function. The proposed analysis yields robust estimates of the portion of genetic variance to the total variance of $y(t)$ in addition to estimates of the underlying genetic and environmental autocorrelation structures. Moreover, the analysis applies in situations where the number T of repeated measures exceeds the number $4N$ of subjects. In such cases $\mathbf{S}_y$ is singular, but its decomposition into $\mathbf{S}_g$ and $\mathbf{S}_e$ can still be obtained from standard genetic analysis of projections y^*_i associated with nonzero eigenvalues. In general, the estimated autocorrelation structures given by $\hat{\mathbf{S}}_g$ and $\hat{\mathbf{S}}_e$ enable a further characterization of the genetic and environmental processes in terms of parametric dynamic models. These results can be amplified and generalized in several ways, some of which are discussed shortly. But first we turn to a concise presentation of parametric models for $G(t)$ and $E(t)$ and then proceed with a few illustrative applications of the proposed analysis.

PARAMETRIC MODELING

One of the outcomes of the proposed analysis is a pair of $T \times T$ matrices $\hat{\mathbf{S}}_g$ and $\hat{\mathbf{S}}_e$ describing the covariance function of $G(t)$ and $E(t)$, respectively. Up to this point it is immaterial to the proposed genetic

analysis whether these covariance functions are stationary or nonstationary. Consequently, these covariance functions can be used in order to obtain estimates of the time course of $G(t)$ and $E(t)$, $t = 1, \ldots, T$ [so-called Wiener filtering (cf. Ahmed and Rao, 1975)]. On the other hand, $\hat{\mathbf{S}}_g$ (similar remarks pertain to $\hat{\mathbf{S}}_e$) can serve as a starting point for the identification of a parametric model for $G(t)$. The identification of such a parametric model yields a minimal description of the dynamic process underlying $G(t)$ and thus constitutes a much more economical and interpretable representation than $\hat{\mathbf{S}}_g$ itself (remember that T can be quite large).

A complete description of the identification of parametric time-series models cannot be given within the scope of this article. The interested reader is refered to the substantial literature on this subject (e.g., Box and Jenkins, 1976; Kashyap and Rao, 1976). We discuss only some basic steps in the identification of a parametric model for $G(t)$ [again, similar remarks pertain to parametric modeling of $E(t)$]. First, one has to ascertain whether the parameters in a model for $G(t)$ are time-varying or constant. We do not consider models with time-varying parameters but, for the sake of clarity, restrict ourselves to a consideration of an important subset of time-series models with constant parameters. This subset is characterized by stationarity of the covariance function. Remember that the covariance function of $G(t)$ is stationary if

$$c_g(t, t') = c_g(u) = c_g(-u),$$

where $u = t' - t$. If $\mathbf{S}_g$ is stationary, then its expected pattern Σ_g is given by

$$\Sigma_g = \begin{bmatrix} c_g(0) & & & \\ c_g(1) & c_g(0) & & \\ c_g(2) & c_g(1) & c_g(0) & \\ c_g(3) & c_g(2) & c_g(1) & c_g(0) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}.$$

Stationarity of $\hat{\mathbf{S}}_g$ can be tested by

$$\chi^2 = [(4N - 1) - 1/6(2T + 1 - 2/(T + 1))] * [\ln | \hat{\Sigma}_g | - \ln | \hat{\mathbf{S}}_g | + \operatorname{tr}(\hat{\Sigma}_g^{-1} \hat{\mathbf{S}}_g) - T],$$

$$\hat{\Sigma}_g = \{\hat{c}_g(u);\ t, t' = 1, \ldots, T\},$$

$$\hat{\mathbf{S}}_g = \{\hat{c}_g(t, t');\ t, t' = 1, \ldots, T\},$$

$$\hat{c}_g(u) = 1/(T - u) \sum_{t=u+1}^{T} \hat{c}_g(t, t - u),$$

where $4N$ is the number of individuals. As $4N \rightarrow \infty$, χ^2 has a distribution with $T(T-1)/2$ degrees of freedom (Morrison, 1976, p. 248).

If $\hat{\mathbf{S}}_g$ is stationary, then one suitable type of time-series model is given by the general pth-order autoregressive model:

$$G(t) = \sum_{u=1}^{p} \beta(u)G(t-u) + \gamma(t),$$

where $c_\gamma(u) = 0$ if $u \neq 0$, i.e., $\gamma(t)$ lacks autocorrelation. It can be shown (cf. Box and Jenkins, 1976) that the covariance function $c_g(u)$, $u = 0, 1, \ldots, T-1$, is sufficient for the identification of the order p and the determination of initial estimates of the parameters $\beta(u)$, $u = 1, \ldots, p$, and the variance $c_\gamma(0)$ of $\gamma(t)$.

In conclusion, then, the above remarks indicate that a dynamic genetic analysis can be supplemented with a stationarity test of $\hat{\mathbf{S}}_g$ and $\hat{\mathbf{S}}_e$. In case stationarity holds, a pth-order autoregressive model can be identified which yields an economical and interpretable description of the covariance function in question. These supplementary steps can be implanted in a fully algorithmic procedure.

APPLICATIONS TO SIMULATED DATA

In this section we present a few illustrative applications of the proposed analysis. A computer program[4] has been written that (1) generates simulated data according to the basic genetic model for time series, (2) carries out the dynamic genetic analysis based on the Karhunen-Loève expansion, and (3) determines stationarity tests and autoregressive models. In the following we concentrate upon the generation of simulated data and the dynamic genetic analysis. We do not discuss stationarity tests and consider only stationary first-order autoregressive models for $G(t)$ and $E(t)$. Notice, however, that these restrictions are not inherent to the proposed analysis.

For the basic genetic model

$$y_{kmn}(t) = G_{kmn}(t) + E_{kmn}(t),$$

it holds that across MZ twins (i.e., $k = 1$)

$$\text{cor}\,[G_{11n}(t), G_{12n}(t)] = 1, \qquad t = 1, \ldots, T,$$

whereas across DZ twins (i.e., $k = 2$)

[4] On request, a listing of the computer program that generates the data series and carries out the principal-components analysis can be obtained from the authors.

$$\text{cor}\,[G_{21n}(t), G_{22n}(t)] = 0.5, \qquad t = 1, \ldots T.$$

In addition, for $t = 1, \ldots, T$,

$$\text{cor}\,[E_{kmn}(t), E_{k'm'n'}(t)] = 0, \qquad k \neq k', m \neq m', \text{ or } n \neq n'.$$

With these provisions a data set can be generated if process models for $G(t)$ and $E(t)$ have been specified. For illustrative purposes we choose first-order autoregressive models for $G(t)$ and $E(t)$:

$$\begin{aligned} G(t) &= \beta(g)G(t-1) + \gamma(t), \\ E(t) &= \beta(e)E(t-1) + \epsilon(t), \\ c_\gamma(t, t') &= \delta(t'-t)\,\sigma_\gamma^2, \\ c_\epsilon(t, t') &= \delta(t'-t)\,\sigma_\epsilon^2, \end{aligned}$$

where $\delta(\cdot)$ is the Kronecker delta. Letting $|t'-t| = u$, it then follows (cf. Box and Jenkins, 1976) that

$$\begin{aligned} c_g(t, t') &= c_g(0, u) = c_g(u) = \beta^u(g)c_g(0), \\ c_g(0) &= \sigma_\gamma^2/[1-\beta^2(g)], \\ c_e(t, t') &= c_e(0, u) = c_e(u) = \beta^u(e)c_e(0), \\ c_e(0) &= \sigma_\epsilon^2/[1-\beta^2(e)]. \end{aligned}$$

Accordingly, $G(t)$ and $E(t)$ are completely specified by fixing $\beta(g)$, $\beta(e) \in [-1, 1]$ and σ_γ^2, $\sigma_\epsilon^2 \in [0, \infty)$, respectively. Notice that

$$\begin{aligned} c_g(1)/c_g(0) &= \beta(g)c_g(0)/c_g(0) = \beta(g), \\ c_e(1)/c_e(0) &= \beta(e)c_e(0)/c_e(0) = \beta(e). \end{aligned}$$

In addition, notice that for $\beta(g) \neq 0$ the covariance function of $G(t)$ is a decreasing function of lag u, whence $\mathbf{S}_g$ lacks compound symmetry. Also, if $\beta(e) \neq 0$, then $c_e(u)$ is a decreasing function of lag u, and consequently $\mathbf{S}_e$ lacks compound symmetry.

The *IMSL Library* (IMSL, Inc., 1979) contains a useful subroutine (FTGEN) that generates time series according to various process models. This subroutine was used for the simulation of a data set according to the above specifications which can be summarized as follows:

(1) Choose N and T;
(2) choose h^2;
(3) choose $\beta(g)$ and $\beta(e)$;
(4) choose σ_γ^2 and compute σ_ϵ^2 such that $h^2 = c_g(0)/[c_g(0) + c_e(0)]$ given by step 2 obtains;

(5) for each pair of MZ twins $n = 1, \ldots, N$ generate, respectively, $G_{1.n}(t)$, $E_{11n}(t)$ and $E_{12n}(t)$, $t = 1, \ldots, T$, with parameter specifications given by steps 3 and 4, then compute $y_{1mn}(t) = G_{1.n}(t) + E_{1mn}(t)$, $m = 1, 2$;

(6) for each pair of DZ twins $n = 1, \ldots, N$, for $t = 1, \ldots, T$ and $m = 1, 2$ and with parameter specifications given by steps 3 and 4, generate $G_{2mn}(t)$ under the restriction that cor $[G_{21n}(t), G_{22n}(t)] = 0.5$, generate $E_{2mn}(t)$, and compute $y_{2mn}(t) = G_{2mn}(t) + E_{2mn}(t)$.

Four distinct types of data are considered:

(I) $\beta(g) = 0$ and $\beta(e) = 0$; consequently, an observed series $y(t)$ lacks autocorrelation, i.e., $c_y(u) = 0$ if $u \neq 0$ and $\mathbf{S}_y$ has compound symmetry because it is a diagonal matrix;

(II) $\beta(g) = 0.75$ and $\beta(e) = 0$; $y(t)$ now has an intricate pattern of autocorrelation (cf. Granger and Morris, 1976);

(III) $\beta(g) = 0$ and $\beta(e) = 0.75$; again, $y(t)$ has an intricate pattern of autocorrelation; and

(IV) $\beta(g) = 0.75$ and $\beta(e) = 0.75$; the pattern of autocorrelation of $y(t)$ is still more involved than with types II and III.

A type IV data set [i.e., $\beta(g) = \beta(e) = 0.75$] has been generated with $N = 100$, $T = 10$, and $h^2 = 0.5$. $c_\gamma(0)$ and $c_\epsilon(0)$ have been chosen in such a way that $c_g(0) = c_e(0) = 100$. This data set has been analyzed according to the proposed method. First, we consider the maximum-likelihood parameters in the structural models for the mean squares of y^*_i (i.e., the projections on each eigenvector $\hat{\mathbf{p}}_i$, $i = 1, \ldots, 10$). Estimates of σ_g and σ_e corresponding to genetic and environmental influences are given in Table I. Combined χ^2–goodness of fit = 8.70 (df = 20, i.e., 2 df for each y^*_i). These parameter estimates enable computation of the portion of genetic variance in the total variance of $y(t)$: $\hat{h}^2 = 0.47$. Second, estimates of the underlying genetic and environmental correlation functions $\hat{\mathbf{S}}_g$ and $\hat{\mathbf{S}}_e$ are determined. These estimates are shown in Table II. As a last step first-order autoregressive models were fitted to $\hat{\mathbf{S}}_g$ and $\hat{\mathbf{S}}_e$, yielding estimates of $\beta(g)$ and $\beta(e)$. The resulting average estimates were $\beta(g) = 0.58$ and $\beta(e) = 0.66$. In view of the rather short length $T = 10$ of the simulated time series, these results seem quite satisfactory.

Next, letting $N = 100$, $T = 50$, and $h^2 = 0.5$, types I, II, III, and IV data sets have been generated and analyzed according to the proposed method. The results thus obtained are shown in Table III. For reasons of conciseness, estimates of S_g (50 × 50) and S_e (50 × 50) are not shown. Instead, we computed

$$\bar{c}_g(0) = 1/T \sum_{t=1}^{T} c_g(t, t) \qquad \text{and}$$

Table I. Maximum-Likelihood Estimates of Genetic and Environmental Standard Deviations for Projections on Each Eigenvector $\hat{\mathbf{p}}_i$, $i = 1, \ldots, 10$

	σ_g	σ_e
1	4.250	4.736
2	5.083	4.684
3	5.343	5.642
4	6.067	5.598
5	6.308	6.410
6	8.666	6.828
7	8.577	9.318
8	12.412	12.024
9	14.890	14.706
10	16.038	20.510

Table II. Genetic and Environmental Correlation Matrices

Genetic correlations

1									
0.588	1								
0.281	0.610	1							
0.202	0.281	0.518	1						
0.174	0.121	0.301	0.586	1					
0.041	0.036	0.132	0.316	0.572	1				
0.068	0.059	0.134	0.193	0.329	0.591	1			
0.028	0.038	0.053	0.010	0.111	0.265	0.591	1		
0.056	0.033	0.034	0.034	0.014	0.145	0.289	0.596	1	
0.016	0.008	0.059	0.035	0.009	0.065	0.197	0.417	0.630	1

Environmmental correlations

1									
0.653	1								
0.463	0.672	1							
0.289	0.370	0.638	1						
0.222	0.296	0.500	0.681	1					
0.076	0.132	0.335	0.419	0.618	1				
0.105	0.127	0.282	0.300	0.464	0.661	1			
0.139	0.179	0.199	0.148	0.291	0.413	0.655	1		
0.070	0.147	0.192	0.160	0.200	0.287	0.424	0.634	1	
0.038	0.131	0.137	0.137	0.165	0.197	0.291	0.480	0.698	1

Table III. True and Estimated Parameter Values for Types I–IV Data Series

	True					Estimated				
	h^2	$\beta(g)$	$\beta(e)$	σ_γ^2	σ_ϵ^2	$\hat{h}^2$	$\hat{\beta}(g)$	$\hat{\beta}(e)$	$\hat{\sigma}_\gamma^2$	$\hat{\sigma}_\epsilon^2$
I	0.5	0	0	1	1	0.482	−0.040	−0.016	0.966	1.040
II	0.5	0.75	0	0.438	1	0.476	0.640	0.065	0.570	1.057
III	0.5	0	0.75	1	0.438	0.527	0.104	0.569	1.045	0.640
IV	0.5	0.75	0.75	0.438	0.438	0.511	0.692	0.658	0.531	0.533

$$\bar{c}_g(1) = 1/T - 1 \sum_{t=2}^{T} c_g(t, t-1), \qquad \text{whence}$$

$$\hat{\beta}(g) = \bar{c}_g(1)/\bar{c}_g(0) \qquad \text{and}$$

$$\hat{\sigma}_\gamma^2 = [1 - \beta^2(g)]\bar{c}_g(0).$$

Similar computations yield $\hat{\beta}(e)$ and $\hat{\sigma}_\epsilon^2$.

Inspection of Table III shows that the proposed analysis manages to recover the underlying genetic structure of the time series. Consider, for instance, the results obtained with type III data. Here, the observed series $y(t)$ have substantial autocorrelation, whereas the genetic series are uncorrelated across time. Consequently, it would seem difficult to identify correctly the lack of genetic autocorrelation, yet our method succeeds in doing so. The same remarks can be made with respect to type II data, where the observed series $y(t)$ have considerable autocorrelation, but the environmental series are uncorrelated across time. In conclusion, then, the proposed method would seem to constitute a viable approach to the genetic analysis of time series.

DISCUSSION

The Karhunen-Loève transformation constitutes a descriptive approach to time-series analysis. It involves a general one-to-one mapping as a means to ease the genetic analysis under consideration. In contrast, a simplex analysis of longitudinal data (Boomsma and Molenaar, 1987) hinges upon the choice of a particular time-series model and thus constitutes a modeling approach. This implies that any misspecifications of the time-series model in question will lead to erroneous results, whereas no such errors can arise in a Karhunen-Loève analysis. Of course, the increased generality of the latter approach is gained at the cost of a decrease in power. In a sense, this state of affairs is analogous to the distinction between parametric and nonparametric approaches to statistical

analysis [cf. Jenkins and Watts (1968) for a similar point of view]. Only after a dynamic genetic analysis based upon the Karhunen-Loève expansion has been carried out can time-series models be fitted to the obtained genetic and environmental covariance functions. These supplementary steps leave the nonparametric character of the Karhunen-Loève analysis intact.

Notwithstanding the intended nonparametric character of the present approach, the standard genetic analysis carried out with respect to projections y^*_i on each eigenvector yields both appropriate likelihood-ratio tests and standard errors of the estimated genetic and environmental parameters. Furthermore, standard errors of the estimated eigenvalues and eigenvectors of $\mathbf{S}_y$ can be regularly obtained (cf. Morrison, 1976). However, $\hat{h}^2$, $\hat{e}^2$, $\hat{\mathbf{S}}_g$, and $\hat{\mathbf{S}}_e$ are derived from these original estimates by means of (inverse) transformation, and at present, the associated standard errors are unknown. We plan to address this issue in the near future.

The Karhunen-Loève transformation has been described for the case in which each observed series has a constant mean function. Although these examples may have some psychophysiological relevance, it is important to note that the proposed genetic analysis remains valid if the observed time series has a time-varying mean function. Obviously, one should then take $\mathbf{S}_y$ to be the matrix of second-order moments about zero. No additional principles are being involved save for the rather strong requirement that the time-varying mean function is invariant across different subjects.

If it is assumed that the covariance function of the observed series is stationary, then $\mathbf{S}_y$ has a particularly regular form and is called a Toeplitz matrix (Brillinger, 1975). The assumption of a stationary covariance function can be tested as described earlier. The Karhunen-Loève transformation of a sufficiently long covariance stationary time series converges to the discrete Fourier transformation (Brillinger, 1975). Stated otherwise, the eigenvectors of a high-dimensional Toeplitz matrix converge to the Fourier axes. This result has enormous computational advantage, as the Fourier axes are analytically given and a numerical determination of the eigenvalue decomposition of $\mathbf{S}_y$ becomes superfluous. In addition, the discrete Fourier transformation has several additional advantages which relate to its robustness against time shifts of, e.g., the underlying genetic and environmental series. However, in view of the various technical intricacies inherent to a quantitative genetic spectral analysis of time series, this approach will be elaborated in a separate publication.

Until now we have considered the genetic analysis of univariate time series. The Karhunen-Loève transformation cannot be generalized to multivariate time series because the sets of eigenvectors obtained with at

least three matrices of auto- and cross-covariances generally will be different. Thus, it is impossible to arrive at a single space in which each component series of a multivariate time series has uncorrelated projections on the base vectors. Instead, one could proceed with a generalized Karhunen-Loève transformation in which a reduced set of base vectors with required properties is constructed according to a recursive procedure (Molenaar, 1981; Stobberingh, 1972). On the other hand, one could invoke dynamic factor analysis of multivariate time series (Molenaar, 1985) or discrete Fourier transformation leading to complex-valued spectral analysis. The latter approach is quite appealing, as it allows for frequency-dependent structural modeling in a way that resembles the usual genetic covariance models proposed by, e.g., Martin and Eaves (1977).

REFERENCES

Ahmed, N., and Rao, K. R. (1975). *Orthogonal Transforms for Digital Signal Processing*, Springer-Verlag, New York.

Boomsma, D. I., and Molenaar, P. C. M. (1987). The genetic analysis of repeated measures. I: Simplex models. *Behav. Genet.* **17:** 111–123.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems. II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Ann. Math. Stat.* **25:**484–498.

Box, G. E. P., and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco.

Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory*, Holt, Rinehart & Winston, New York.

Granger, C. W. J., and Morris, M. J. (1976). Time series modeling and interpretation. *J. Roy. Stat. Soc. A* **139:**246–257.

IMSL, Inc. (1979). *IMSL Library Reference Manual Edition 7*, IMSL, Inc., Houston, Tex.

Jenkins, G. M., and Watts, D. G. (1968). *Spectrum Analysis and Its Applications*, Holden-Day, San Francisco.

Kashyap, R. L., and Rao, A. R. (1976). *Dynamic Stochastic Models from Empirical Data*, Academic Press, New York.

Lutzenberger, W., Elbert T., Rockstroh, B., and Birbaumer, N. (1980). Biofeedback of slow cortical potentials. II. Analysis of single event-related potentials by time-series analysis. *Electroenceph. Clin. Neurophysiol.* **48:**302–311.

Martin, N. G., and Eaves, L. J. (1977). The genetical analysis of covariance structure. *Heredity* **38:**79–95.

Molenaar, P. C. M. (1981). *Dynamische factormodellen*, doctoral dissertation, State University of Utrecht, Utrecht (in Dutch).

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika* **50:**181–202.

Morrison, D. F. (1976). *Multivariate Statistical Methods*, McGraw–Hill Kogakusha, Tokyo.

Stobberingh, R. (1972). *Dynamische componenten analyse, een intergratie van componenten- en tijdreeksanalyse*, doctoral dissertation, Catholic University, Tilburg (in Dutch).

Wastell, D. G. (1981). On the correlated nature of evoked brain activity: Biophysical and statistical considerations. In Gaillard, A. W. K., and Ritter, E. (eds.), *Tutorials in ERP Research: Endogenous Components*, North-Holland, Amsterdam.

Edited by C. R. Cloninger